Mygdalis, V., Iosifidis, A., Tefas, A., & Pitas, I. (2014). *Supervised video summarization by content-based video segment selection*. Paper presented at European Conference on Visual Media Production (CVMP), London, United Kingdom.

Supervised video summarization by content-based video segment selection

V. Mygdalis, A. Iosifidis, A. Tefas, I. Pitas

Department of Informatics, Aristotle University of Thessaloniki,

54124, Thessaloniki, Greece

{tefas,pitas}@aiia.csd.auth.gr

Keywords: One class classification, Subclass One-Class SVM, Supervised Video Summarization.

Abstract

In this paper, we formulate the video summarization problem as the one of automatic video segment selection based on one-class classification. We introduce a novel variant of the One-Class Support Vector Machine classifier that exploits subclass information in its optimization problem, in order to jointly minimize the data dispersion within each subclass and determine the optimal decision function. We evaluate the proposed approach in Hollywood movies, where the performance of the proposed SOC-SVM algorithm is compared with that of the OC-SVM.

1 Introduction

In this paper, we describe a video summarization technique that selects a number of video segments from a full length video stream, to formulate a video skim of smaller duration. For each video segment, we employ an activity-based description [2], since most of the available video content has been recorded in order to capture human activity. Subsequently, by exploiting salient video segment paradigms (movie trailers in the case of Hollywood movies), we formulate an one class classification scheme, since a definition of negative classes is difficult and subjective. In order to enhance the classification performance, we extend the One-Class SVM formulation [1], so that to exploit subclass information appearing in the salient video segment paradigms.

2 Subclass One-Class Support Vector Machine

The Subclass One-Class SVM (SOC-SVM) algorithm exploits the subclass information appearing in the salient video segments and is expressed by:

$$\mathbf{S} = \sum_{k=1}^{K} \sum_{\mathbf{z}_i \in \mathcal{M}_k} \frac{N_k}{N} (\mathbf{z}_i - \mathbf{m}_k)(\mathbf{z}_i - \mathbf{m}_k)^T ,$$

where $\mathbf{z}_i$, i=1,…,N are the vectorial representations of the training samples and $\mathbf{m}_k$ is the mean vector of subclass k (formed by N_k samples). Subsequently, the following optimization problem is solved:

$$\begin{aligned} \min_{\mathbf{w},\xi_i,\rho} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{S}\mathbf{w} + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho \\ \textbf{s.t.:} \quad & \mathbf{w}^T\mathbf{z}_i \geq \rho - \xi_i,\ i = 1,\dots,N, \\ & \xi_i \geq 0, \end{aligned}$$

in order to determine the optimal vector **w** that can be used to classify test samples as samples belonging to the training class or not. **w** is given by:

$$\mathbf{w} = \mathbf{S}^{-1}\sum_{i=1}^{N}\alpha_i\mathbf{z}_i .$$

3 Experimental results

We have applied the method to three, full length Hollywood movies belonging to action, adventure and drama categories, respectively. For training the One-Class classifiers we have employed the trailers of eighteen Hollywood movies belonging to action, comedy, thriller and drama categories. Resulted summaries include *p*% video segments of a movie, and are rated with a precision measurement respective to a ground truth selection of video segments which have been employed to form the actual movie trailer. Experimental results are shown in Table 1.

***p* %**	**10**	**20**	**30**	**40**	***50***
OC - SVM	11,68	22.39	33.53	41.36	53.96
SOC – SVM	**21,19**	**29.11**	**38.37**	**54.62**	**62.68**

Table 1 : Mean performance for different summary lengths.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART).

References

[1] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson. “Estimating the support of a high-dimensional distribution”, Neural Computation, vol. 13, pp 1443–1471, (2001)

[2] H. Wang, A. Klaser, C. Schmid, and C. L. Liu. “Dense trajectories and motion boundary descriptors for action recognition”, International Journal of Computer Vision, vol. 103, pp.60–79, (2012)